United States
Department of
Agriculture

National
Agricultural
Statistics
Service

Research Division

SRB Research Report
Number SRB-93-01

March 1993

A COMPARISON OF FOUR ALTERNATIVE WEIGHTED ESTIMATORS TO THE OPEN ESTIMATOR FOR USE IN THE AGRICULTURAL LABOR SURVEY

Cheryl L. Turner

A COMPARISON OF FOUR ALTERNATIVE WEIGHTED ESTIMATORS TO THE OPEN ESTIMATOR FOR USE IN THE AGRICULTURAL LABOR SURVEY, by Cheryl L. Turner, Applications Research Section, Survey Research Branch, Research Division, National Agricultural Statistics Service, United States Department of Agriculture, Washington, DC 20250-2000, March 1993, Report No. SRB-93-01.

ABSTRACT

The Agricultural Labor Survey (ALS), conducted by the National Agricultural Statistics Service within the United States Department of Agriculture, is a multiple frame survey consisting of a list sampling frame and the non-overlap portion of the area sampling frame. This study compares four alternative weighted estimators of the peak number of hired workers (the operational, modified weighted, Hanuschak-Keough strata mean, and the Hanuschak-Keough strata median estimates) in the non-overlap domain to the currently used open estimator approach. Historically, the open estimator tends toward a downward bias, and it also has a higher coefficient of variation (CV). Of the four alternatives, the modified weighted is the recommended estimator. The modified weighted estimator improved upon the downward bias of the open estimator and obtained a smaller CV, while at the same time making a more cost effective survey.

KEY WORDS

Agricultural labor survey; Non-overlap; Open estimator; Weighted estimator.

ACKNOWLEDGEMENTS

The author would like to thank Lee Brown for reviewing both the early and final drafts of this paper; I would also like to thank Bill Iwig for his technical advice.

TABLE OF CONTENTS

INTRODUCTION

The National Agricultural Statistics Service (NASS) within the United States Department of Agriculture (USDA) annually conducts a June Agricultural Survey (JAS). The JAS is a multiple frame survey, consisting of both a list frame and an area frame. The area frame is stratified according to land usage or the percent of cultivation. The area frame is further subdivided into overlap (OL) and non-overlap (NOL) domains. The overlap portion of the area frame is composed of farming operations which are also found on the list frame. The non-overlap contains those farming operations which are not found on the list frame.

The JAS begins the survey year and is the largest survey of the year for NASS. Follow-on survey samples are derived from a list sampling frame and a sample of the area frame. The Agricultural Labor Survey (ALS) is a multiple frame follow-on survey. It provides estimates of the number of farm workers and of the wage rates paid to those farm workers. Currently, the non-overlap estimate for the ALS is derived using an open estimator. The ALS open estimator is based on a sample of NOL Resident Farm Operators (RFO's) from forty percent of the area segments used in the JAS. (A segment is a piece of land that is the primary sampling unit in the NASS area frame sampling plan.) By definition, the open estimator excludes all non-Resident Farm Operators. An alternative to the open estimator is a weighted estimator. The weighted estimator is generated from a sample of <u>all</u> NOL farm operators, both RFO and non-RFO. The weighted estimator has historically had a smaller coefficient of variation (CV) than the open estimator because the weighted estimate is generated from a larger group of farm operators.

Four weighted estimators were evaluated for possible use in the ALS. They were the operational, modified weighted (modified), Hanuschak-Keough strata mean (H-K mean), and the Hanuschak-Keough strata median (H-K median). Each weighted estimator was compared against the current open estimator. As stated above, the open estimator tends to have a larger CV due to sampling only from the RFO's. Also, in 1981, Nealon (1) stated that the open estimator tends toward a downward bias. One reason for this downward bias is because RFO's are missed during the prescreening stage of the JAS. In 1988, Bosecker and Clark (2) reiterated the downward bias of the open estimator and also proposed a modified weighted estimator whose results were not significantly different from those of the operational weighted estimator. In light of these previous findings, the aforementioned operational and modified weighted estimates, along with the H-K mean and H-K median, were compared with the open estimator for possible use in the ALS.

This report represents the comparative analysis done on these alternative weighted estimators. All estimators used the "peak number of hired workers" from 1991 JAS data. The JAS area

questionnaire obtains the expected "peak number of hired workers" for the survey year. This number is then used to define the NOL strata for the follow-on ALS. These strata are defined in Appendix A. This study was done independently on both the 17 labor regions and the eleven monthly and seasonal states. (Monthly and seasonal ALS were conducted in 1991 and 1992.) The 17 labor regions contain the 48 contiguous states within the United States. Most of the states within the 17 labor regions conduct a quarterly ALS, with the exception of the eleven monthly and seasonal states. California, Florida, New Mexico, and Texas conduct a monthly ALS, while Michigan, New York, North Carolina, Oregon, Pennsylvania, Washington, and Wisconsin conduct a seasonal ALS. The seasonal survey is conducted in January and then again in April through October.

STUDY DESIGN

Data for this survey were collected during the 1991 JAS and represent the NOL domain. The item of interest was the peak number of hired agricultural workers for the survey year. The data were evaluated at the regional level and at the state level (for the eleven monthly and seasonal states). There are 17 labor regions within the United States. They are defined as follows:

Region	States
Northeast I	Connecticut, Maine, Massachusetts, New Hampshire, New York, Rhode Island, Vermont
Northeast II	Delaware, Maryland, New Jersey, Pennsylvania
Appalachian I	North Carolina, Virginia
Appalachian II	Kentucky, Tennessee, West Virginia
Southeast	Alabama, Georgia, South Carolina
Lake	Michigan, Minnesota, Wisconsin
Cornbelt I	Illinois, Indiana, Ohio
Cornbelt II	Iowa, Missouri
Delta	Arkansas, Louisiana, Mississippi
Northern Plains	Kansas, Nebraska, North Dakota, South Dakota
Southern Plains	Oklahoma, Texas
Mountain I	Idaho, Montana, Wyoming
Mountain II	Colorado, Nevada, Utah
Mountain III	Arizona, New Mexico
Pacific	Oregon, Washington
Florida	Florida
California	California

Florida and California are the only single state regions. This is because their agricultural labor is unique and their data are not easily grouped with any of their neighboring regions.

The eleven monthly and seasonal states are defined as:

MONTHLY	SEASONAL
California	Michigan
Florida	New York
New Mexico	North Carolina
Texas	Oregon
	Pennsylvania
	Washington
	Wisconsin

THE WEIGHTED ESTIMATORS

Two types of estimators were being evaluated, an open and a weighted estimator. For an open estimator, the location of the operator's residence is used to uniquely associate every farm with only one segment. A weight of one is assigned if the tract operator lives within the selected segment (if the tract operator is an RFO), and a weight of zero is assigned otherwise. Conversely, the weighted estimator apportions a farm's activities to a segment by weighing the data relative to the fraction of the farm's acreage that lies within the segment boundary. Therefore, one farm may contribute to the data in several segments.

As stated earlier, the ALS open estimator is based on a sample of NOL RFO's from forty percent of the area segments used in the JAS. In contrast, an ALS weighted estimator would be based on the same sample size being selected from all NOL operations (both RFO's and non-RFO's) from the same forty percent of area segments. The respondents selected using an ALS weighted estimator would have been selected from a larger pool of potential respondents. In sampling from the larger pool of respondents, there is the potential for a reduction in the CV.

The operational, H-K mean, H-K median, and the modified weighted were the four weighted estimators being evaluated.

Operational

The operational weighted estimator is the weighted estimator traditionally used in NASS surveys. It merely assigns an "operational" weight of tract acres divided by total farm acres for each farming operation even partially contained within the segment. (Where the tract acres are the acres residing within a sampled segment.) This estimator prorates farm level data to the segment level.

Hanuschak-Keough strata mean and median

This weighted estimator is similar to the operational weighted estimator, but it attempts to limit potential outliers by controlling the value of the weight. There are occasions when the exact farm acreage is neither obtainable nor known. This happens when the respondent either would not or could not give the correct farm acreage. In these instances the tract acreage and farm acreage may be recorded as equal (plus perhaps a token acre for the farmstead) on the JAS. Although this problem has been recognized and emphasized at training schools, it still exists (but to a lesser degree). Hanuschak and Keough (3) proposed a solution for this specific type of problem. In some cases the equality of the tract and farm acres is accurate. However, if the farm acres should have been substantially larger than the tract acres, the "operational" weight would be nearly or equal to one when it should have been considerably lower. This problem leads to a great overexpansion of the survey data. And conversely, there could be an underexpansion of the survey data if tract acres were underreported.

Hanuschak and Keough recommended a more robust estimator than the standard "operational" weight. A robust estimator is relatively insensitive to slight departures from the assumptions of normality. The Hanuschak-Keough estimators replaced the "operational" weight with a robust weight for all NOL tracts (or observations) in which someone other than the operator or the operator's spouse responded. The Hanuschak-Keough estimators will guard against large overexpansions or underexpansions of the survey data. Consider the following respondent codes as defined in the JAS survey:

Respondent Code
1 = Operator/Manager
2 = Spouse
3 = Other
4 = Observed Refusal
5 = Observed Non-refusal

The Hanuschak-Keough estimators replaced the "operational" weight for all NOL observations containing respondent code 3, 4, or 5 with a more robust weight. Within each land use strata, the Hanuschak-Keough strata mean estimator replaced the denominator of the "operational" weight for those observations containing respondent codes 3, 4, or 5 with the average farm acreage from the respondent code 1 and 2 observations. The Hanuschak-Keough strata median estimator replaced the denominator within each land use strata for those same observations with the median farm acreage from the respondent code 1 and 2 observations. For example, a tract contained in strata 11 with respondent code 3 will have the following weights:

"operational" weight = $\frac{\text{NOL tract acreage}}{\text{farm acreage}}$

Hanuschak-Keough strata mean weight

$$= \frac{\text{NOL tract acreage}}{\text{average farm acreage for NOL strata 11 observations with respondent code 1 or 2}}$$

Hanuschak-Keough strata median weight

$$= \frac{\text{NOL tract acreage}}{\text{median farm acreage for NOL strata 11 observations with respondent code 1 or 2}}$$

Modified Weighted

The modified estimator was originally proposed by Bosecker and Clark. It is an effort to eliminate screening for farm operators in densely populated segments. In reducing the amount of survey screening, the cost of conducting the survey is greatly reduced.

The modified estimator is especially suited to the measurement of rare populations, and the number of farm operators among the general population (particularly in residential areas) certainly qualifies as rare. The modified weighted estimator will exclude up to one half acre for non-agricultural land devoted to residential purposes (such as the house and yard). For residential agricultural tracts, the residential area would be subtracted from the weight's numerator and denominator; for non-resident agricultural tracts, the residential area would be subtracted just from the weight's denominator. Since the modified weight would be zero for small tracts consisting of only a house and yard, screening for farm operators in residential areas would be unnecessary.

The modified weight assumed 1/2 acre for all residences, except where it was known that the farmstead was less than 1/2 acre. To obtain the modified weights, the following calculations were performed:

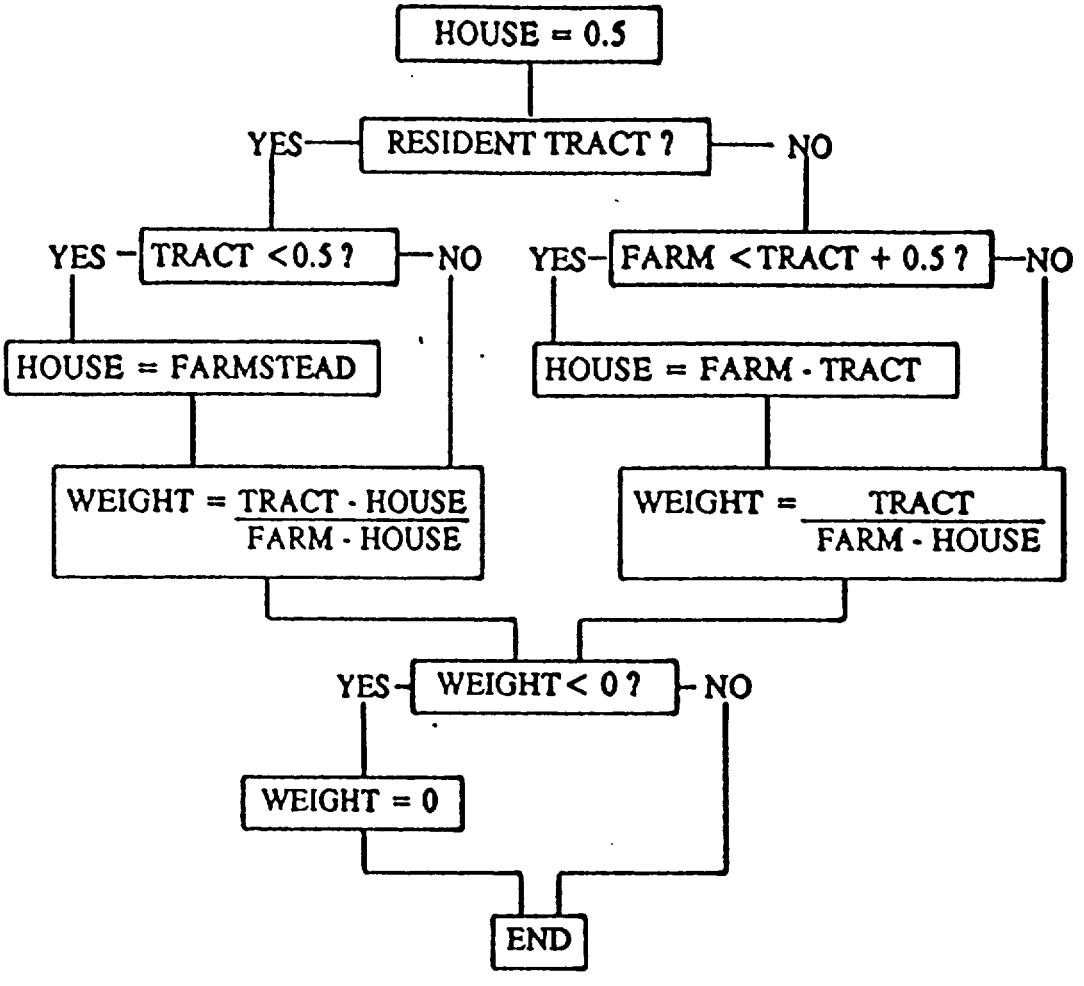
modified weight for residents, $\frac{\text{total tract area - 1/2}}{\text{total farm area - 1/2}}$

modified weight for non-residents, $\frac{\text{total tract area}}{\text{total farm area - 1/2}}$

The exact flow diagram for the modified weights is given in Table 1 on the following page. In the table, "HOUSE" indicates the house and yard acreage, "FARMSTEAD" indicates the farmstead acreage, "TRACT" indicates the tract acreage, and "FARM" indicates the total farm acreage.

The expanded peak number of hired workers was calculated using the open estimator and each of the alternative weighted estimators. These numbers are reported at both the regional and state levels in Appendix B. The corresponding CV's for these estimates were also calculated at the regional and state levels, and they are reported in Appendix C.

TABLE 1: The decision diagram used in calculating the modified weight.



ANALYSIS

NOL estimates were generated for the peak number of hired workers. Both the open and the weighted estimators were generated using the same number of tracts and the same tract information. Identical analyses were used to independently compare each of the four alternative estimates with the current open estimate of the peak number of hired workers. The formula for calculating these five estimates and their corresponding variances is contained in Appendix D. Univariate paired t-tests were conducted at the regional level for the 17 regions and at the state level for the eleven monthly and seasonal states on each alternative estimator versus the open estimator. Appendix E contains a detailed description of the univariate paired t-tests. These t-tests will determine if the alternative estimate was significantly different from the open estimate. The paired t-test will test the following hypotheses for each alternative estimate:

H_O: $Y_{diff} = 0$ versus H_A: $Y_{diff} <> 0$

where Y_{diff} = alternative estimate - open estimate

RESULTS

Univariate paired t-tests were performed on the variable peak number of hired workers. T statistics were calculated for both the 17 labor regions and the eleven monthly and seasonal states for each for the following four scenarios:

1) Operational estimate vs. Open estimate
2) H-K mean estimate vs. Open estimate
3) H-K median estimate vs. Open estimate
4) Modified estimate vs. Open estimate

Labor Region Results

The test results indicated that most of the comparisons yielded insignificant differences (alpha = .05) at the regional level. Therefore, there were negligible differences between each of the four alternative estimators and the open estimator for these regions.

The test results also indicated that some significant differences (alpha = .05) did exist at the regional level. Significant differences between each of the four alternative estimates and the open estimate existed in the Delta region and the Southern Plains region. In the Appalachian II region and the Southeast region,

significant differences existed for all comparisons but the H-K mean estimate and the open estimate. Significant differences existed in the Pacific region between each the operational and modified estimates and the open estimate. And lastly, the Northern Plains and California regions obtained significant differences between the H-K median estimate and the open estimate. Table 2 contains the 17 regions and their significance level for each of the four comparisons.

TABLE 2: Significance levels from the univariate t-tests comparing each of the alternative estimates (operational estimate, H-K mean estimate, H-K median estimate, and modified estimate) versus the open estimate at the regional level.

	SIGNIFICANCE LEVELS			
REGION	Open vs. Operational	H-K mean	H-K median	Modified
Northeast I	0.27908	0.48406	0.32240	0.26420
Northeast II	0.22972	0.23408	0.25908	0.22972
Appalachian I	0.38666	0.26696	0.44780	0.39228
Appalachian II	0.04170*	0.09350	0.01922*	0.03850*
Southeast	0.04980*	0.05892	0.00848*	0.04788*
Lake	0.58424	0.51166	0.40886	0.58502
Cornbelt I	0.13748	0.58284	0.16368	0.13746
Cornbelt II	0.71026	0.67538	0.59166	0.70768
Delta	0.00014*	0.00006*	0.00000*	0.00014*
Northern Plains	0.20848	0.19248	0.03574*	0.21208
Southern Plains	0.00290*	0.00092*	0.00014*	0.00288*
Mountain I	0.49838	0.50186	0.53880	0.49846
Mountain II	0.91686	0.93648	0.34424	0.92900
Mountain III	0.81320	0.95022	0.82936	0.79740
Pacific	0.03932*	0.14166	0.14340	0.03942*
Florida	0.66308	0.37448	0.26044	0.65472
California	0.49140	0.75274	0.00074*	0.49340

where * indicates a significant difference at alpha = .05

As stated above, both the Delta and Southern Plains regions obtained significantly different results for the four alternative estimators as compared to the open estimate. Table 3 contains further examination of these two regions:

Table 3: Significance levels of each of the alternative estimates (operational estimate, H-K mean estimate, H-K median estimate, and modified estimate) versus the open estimate at the state level for the Delta and Southern Plains regions.

	SIGNIFICANCE LEVELS			
	Open vs. Operational	H-K mean	H-K median	Modified
Delta Region				
Arkansas	0.00666*	0.00430*	0.00252*	0.00662*
Louisiana	0.00104*	0.00186*	0.00036*	0.00102*
Mississippi	0.39874	0.23008	0.12848	0.40106
Southern Plains Region				
Oklahoma	0.69204	0.75822	0.41502	0.68854
Texas	0.00258*	0.00060*	0.00016*	0.00260*

where * indicates a significant difference at alpha = .05

As shown in Table 3, Arkansas, Louisiana, and Texas were the dominating states within their respective regions. All states were significantly different with respect to the alternative estimate vs. the open estimate. When Arkansas, Louisiana, and Texas were evaluated individually, one tract often accounted for the majority of difference between the alternative estimates and the open estimate. For example, within Texas there was one tract which produced the following estimates for the expanded peak number of hired workers for the alternative estimates and the open estimate:

Estimate	Single Tract Expansion	State Expansion	% of State
Open	0.00	25765.69	0.00
Operational	4565.71	52627.95	8.67
H-K mean	2229.61	52918.15	4.21
H-K median	5572.61	62303.48	8.94
Modified	4574.70	52790.61	8.67

where % of State = $\frac{\text{Single Tract Expansion}}{\text{State Expansion}} * 100$

This one tract in Texas made no contribution to the peak number of hired workers for the open estimate. But for each of the four alternative weighted estimates, this tract alone contributed between four and eight percent of Texas' state level expansion for the peak number of hired workers. The differences in these estimates were due in part to the farmer living outside of the selected segment (and therefore having an open weight of 0), while at the same time having a positive number of hired workers.

In following with previous findings, Appendix B shows that the open estimate was the lowest estimate (due to a downward bias) in 12 of the 17 regions, while the H-K median was the highest estimate in 11 of the 17 regions. The operational, H-K mean, and modified estimates were most often found between these two extremes.

Appendix C shows that the CV for the open estimator was the largest CV in 13 of the 17 regions. This supports the notion that sampling from a smaller sample size (only the RFO's) will increase the CV. The CV's for the four weighted estimators were (overall) considerably smaller than those for the open estimator, but none of the alternatives distinguished itself as having the lowest CV.

State Level Results

Mostly insignificant differences (alpha = .05) also existed at the state level. And as with the regional level results, this indicated that there were negligible differences between each of the four alternative estimators and the open estimator for the monthly and seasonal states.

The test results at the state level also indicated that some significant differences (alpha = .05) did exist. Significant differences between all four of the alternative estimates and the open estimate existed only in Texas (as was illustrated above). There were significant differences in Washington between the operational estimate and the open estimate and also between the modified weighted estimate and the open estimate. A significant difference also existed between the H-K median estimate and the open estimate in California. Table 4 contains the significance levels of the eleven monthly and seasonal states for each of the four comparisons.

TABLE 4: Significance levels from the univariate t-tests comparing each of the alternative estimates (operational estimate, H-K mean estimate, H-K median estimate, and modified estimate) versus the open estimate at the state level for the eleven monthly and seasonal states.

	SIGNIFICANCE LEVELS			
STATE - MONTHLY	Open vs. Operational	H-K mean	H-K median	Modified
California	0.49140	0.75274	0.00074*	0.49340
Florida	0.66308	0.37448	0.26044	0.65472
New Mexico	0.24772	0.38356	0.30054	0.23390
Texas	0.00258*	0.00060*	0.00016*	0.00260*
STATE - SEASONAL				
Michigan	0.18318	0.18032	0.14718	0.18554
New York	0.10822	0.28738	0.15140	0.10754
North Carolina	0.37800	0.36648	0.52592	0.38294
Oregon	0.66710	0.60028	0.60102	0.66460
Pennsylvania	0.22336	0.23320	0.24044	0.22310
Washington	0.02902*	0.13366	0.15002	0.02904*
Wisconsin	0.12258	0.14206	0.14276	0.12306

where * indicates a significant difference at alpha = .05

Also, as with the regional results, Appendix B shows that the estimates were lowest for the open estimator in 7 of the 11 states and the estimates were highest for the H-K median estimator in 8 of the 11 states. The operational, H-K mean, and modified estimators were barely distinguishable from each other, each lying between the two extremes. Appendix C again shows the open estimator CV as the largest CV in 7 of the 11 states. The four weighted estimator CV's again obtained smaller CV's than the open CV, while not substantially differing from one another.

CONCLUSIONS AND RECOMMENDATIONS

This paper evaluated four alternative weighted estimators (the operational, Hanuschak-Keough strata mean, Hanuschak-Keough strata median, and the modified operational) of the peak number of hired workers and compared them to the current open estimator approach. These evaluations were made at both the labor region and state level. When considering only the estimates and their corresponding CV's, it was evident that the open estimate was biased downward, while at the same time having an increased CV. This indicated that there was a need for a "better" estimator with a smaller CV.

The analyses indicated that, for the most part, insignificant differences existed between the open estimator and any of the four alternative weighted estimators. However, significant differences were also found. The Delta and Southern Plains regions were both significantly different for all four comparisons. Further review of these two regions indicated that one state within the region was primarily responsible for the significant difference. And, in reviewing that state, one (or several) tracts accounted for a substantial percentage of the estimation difference. This indicated that one (or several) tracts within a state could make a region (or state) significantly different.

When there was no significant difference between the alternative and the open estimate, any of the weighted estimators could be considered as a viable selection. Each of the alternative weighted estimators has a smaller CV than the open estimator. But the H-K median estimator also has a strong upward bias, which greatly overestimates the peak number of hired workers. This upward bias negates the H-K median as an adequate alternative to the open estimator. When selecting between the remaining weighted estimators, significant differences were considered. Of the three remaining alternative weighted estimators, more research is recommended on the Hanuschak-Keough strata mean. While the original prognosis on the H-K mean was positive, this is the first study done utilizing this estimator and more positive results are needed before a conclusion can be reached. The operational estimator is a tried and proven estimator. It had a smaller CV than the open estimator and also improved upon the downward bias of the open estimator. But the recommended alternative is the modified weighted. This estimator achieved the accuracy levels of the operational estimate, while also eliminating the JAS screening for farmers in the more densely populated segments, and thus reduced the overall survey cost. More research is also recommended on a combined estimator based on the modified weighted estimator and the H-K mean. This new combined estimator would merge the strong points of both estimators. It would reduce the screening requirements for potential farm operators within residential areas while, at the same time, lessening the effect of any potential outliers.

REFERENCES

[1] Bosecker, Raymond R. and Michael S. Clark. "Modifying the Weighted Estimator to Eliminate Screening Interviews in Residential Areas", U.S. Department of Agriculture, National Agricultural Statistics Service, 1988.

[2] Dillard, Dave and Dale Atkinson. "The Use of JAS Segments and the Commodity Weighted Estimator on FCRS", U.S. Department of Agriculture, National Agricultural Statistics Service, 1989.

[3] Hanuschak, George and Gary Keough. Memo subject "One Outlier Control Method for Immediate Implementation Consideration", April 1991.

[4] McClung, Gretchen. "A Commodity Weighted Estimator", U.S. Department of Agriculture, National Agricultural Statistics Service, 1988.

[5] McClung, Gretchen. "The Commodity Weighted Estimator - An Analysis of 1987 June Agricultural Survey Data", U.S. Department of Agriculture, National Agricultural Statistics Service, 1989.

[6] Nealon, Jack. "An Evaluation of Alternative Weights for a Weighted Estimator", U.S. Department of Agriculture, Statistical Reporting Service, 1981.

[7] Pontius, Robert G. Jr. "Modified Agricultural Weighted Estimators", U.S. Department of Agriculture, National Agricultural Statistics Service, 1990.

[8] U.S. Department of Agriculture (1983): Scope and Methods of the Statistical Reporting Service. Publication No. 1308. Washington, D.C.

Appendix A: Non-Overlap Labor Strata Definitions

LABOR STRATUM	JAS RESPONDENT CODE	DESCRIPTION
9	1, 2, 3	PLF >= 10
8	1, 2, 3	5 <= PLF <= 9
7	1, 2, 3	1 <= PLF <= 4
6	1, 2, 3	PLF = 0 (FCRS sales code >= 6)
5	1, 2, 3, 4, 5	PLF = 0 (FCRS sales code < 6, refusals, inaccessibles)

where,

PLF = Peak Labor Force
FCRS = Farm Costs and Returns Survey

and the JAS Respondent Code is defined as follows:

Respondent Code
1 = Operator/Manager
2 = Spouse
3 = Other
4 = Observed Refusal
5 = Observed Non-refusal

Appendix B: Regional and State level estimates for the peak number of hir workers contained in the nonoverlap sample of the ALS

PEAK NUMBER OF HIRED WORKERS

REGION	Open	Operational	H-K mean	H-K median	Modifi
Northeast I	20694.05	24004.10	22851.72	23667.74	24110.
Northeast II	29380.67	16987.91	17110.17	17714.08	16968.
Appalachian I	90750.63	77050.99	73284.80	78489.32	77146.
Appalachian II	105861.39	126928.61	123413.01	131791.55	127606.
Southeast	34367.01	47059.97	49483.26	55375.85	47210.
Lake	67690.76	76183.25	78396.38	82017.12	76176.
Cornbelt I	31096.45	55784.45	33350.42	37581.48	57693.
Cornbelt II	31602.88	37713.00	38544.85	40530.45	37780.
Delta	29050.10	46066.63	47619.40	50937.16	46101.
Northern Plains	9379.25	13960.28	14185.88	17845.36	13941.
Southern Plains	53472.38	81542.11	81567.29	92812.59	81720.
Mountain I	29312.15	19000.76	19084.06	19942.92	18989.
Mountain II	22055.56	21876.02	22196.05	23740.49	21901.
Mountain III	11184.44	11585.31	11288.53	11548.99	11622.
Pacific	27912.42	51067.92	86808.51	102043.42	51163.
Florida	22930.05	26520.26	32167.40	50985.23	26631.
California	188676.61	116644.80	160427.00	279827.81	116943.

STATE - MONTHLY	Open	Operational	H-K mean	H-K median	Modifi
California	188676.61	116644.80	160427.00	279827.81	116943.
Florida	22930.05	26520.26	32167.40	50985.23	26631.
New Mexico	7253.70	8235.58	7941.93	8090.58	8277.
Texas	25765.69	52627.95	52918.15	62303.48	52790.
STATE - SEASONAL					
Michigan	20420.28	28666.20	28707.60	29698.13	28660.
New York	8226.88	11534.10	10532.10	11160.72	11550.
North Carolina	69224.58	56616.58	56305.63	59879.73	56680.
Oregon	11811.04	10887.20	10686.35	13105.48	10875.
Pennsylvania	22976.61	10476.92	10738.82	10923.35	10450.
Washington	16101.38	40180.71	76122.16	88937.94	40287.
Wisconsin	26293.57	35805.68	37977.40	40607.62	35788.

Appendix C: Regional and State level coefficients of variation (CV's) for the peak number of hired workers contained in the nonoverlap sample of the ALS

CV'S FOR THE PEAK NUMBER OF HIRED WORKERS

REGION	Open	Operational	H-K mean	H-K median	Modified
Northeast I	22.0435	22.6036	22.8876	22.5017	22.6999
Northeast II	38.1151	14.9051	14.5372	14.6197	14.9024
Appalachian I	24.3255	16.8520	16.7761	16.3973	16.8479
Appalachian II	12.8280	10.6717	10.2007	10.3035	10.7249
Southeast	23.7132	14.9524	17.7586	17.2956	14.9250
Lake	29.5622	20.4344	21.0553	21.5770	20.4421
Cornbelt I	19.8291	31.9361	14.7140	14.7716	33.0829
Cornbelt II	41.6895	27.7375	27.3326	26.2352	27.7460
Delta	36.4987	22.7815	22.1617	20.9011	22.7710
Northern Plains	32.4980	18.6555	18.2633	17.0110	18.6961
Southern Plains	20.8959	14.4932	13.5119	13.5024	14.5145
Mountain I	79.4150	44.0498	43.8032	42.0398	43.9821
Mountain II	70.4286	70.4013	69.4025	64.9494	70.3213
Mountain III	34.8162	32.4996	32.8416	32.3386	32.5240
Pacific	31.7965	25.4087	46.6110	49.9226	25.4439
Florida	47.0902	58.5918	52.0934	55.2249	58.4622
California	63.1159	23.5737	24.7544	43.0588	23.6255

STATE - MONTHLY	Open	Operational	H-K mean	H-K median	Modified
California	63.1159	23.5737	24.7544	43.0588	23.6255
Florida	47.0902	58.5918	52.0934	55.2249	58.4622
New Mexico	40.7737	41.4296	42.1560	41.7473	41.4351
Texas	25.7800	15.9966	13.8333	15.0124	16.0628
STATE - SEASONAL					
Michigan	40.0328	33.7366	33.6598	32.8066	33.7762
New York	40.1724	32.3943	32.7684	32.1566	32.3760
North Carolina	29.1263	19.5140	19.7740	19.7912	19.5028
Oregon	35.1417	28.3786	28.7563	25.8208	28.3496
Pennsylvania	48.3601	20.2608	19.5922	19.2956	20.2647
Washington	48.7215	31.3645	53.0010	57.1525	31.3930
Wisconsin	32.6240	31.0134	32.6325	34.1886	31.0126

Appendix D: Formula for calculating the estimate totals and their corresponding variances

A sample estimate of the peak number of hired workers and a variance of that estimate was generated for each of the alternative estimators and for the open estimate. The four alternative estimators being calculated are the operational estimator, Hanuschak-Keough strata mean, Hanuschak-Keough strata median, and modified weighted. The estimates for the state level peak number of hired workers are defined as follows:

$$\hat{Y}_{STATE} = \sum_{l=1}^{s} \sum_{k=1}^{p_l} \sum_{j=1}^{r_{lk}} \hat{Y}'_{lkj}$$

$$= \sum_{l=1}^{s} \sum_{k=1}^{p_l} \sum_{j=1}^{r_{lk}} e_{lkj} \hat{Y}_{lkj}$$

where

s = *the number of land use strata in the state*

p_l = *the number of substrata within land use stratum l*

r_{lk} = *the number of segments within substratum k within land use stratum l*

e_{lkj} = *the expansion factor for segment j, within substratum k, within land use stratum l*

$$\hat{Y}_{lkj} = \sum_{m=1}^{f_{kj}} n_{mlkj} \, w_{mlkj} \, z_{mlkj}$$

where

f_{lkj} = the number of agricultural tracts in segment j, within substratum k, within land use stratum l

n_{mlkj} = the NOL indication for tract m, within segment j, substratum k and land use stratum l
= 1 if tract is NOL, and
= 0 if tract is OL

w_{mlkj} = weight for tract m, within segment j, substratum k and land use stratum l

= tract acres/farm acres, for the operational weighted estimate

= the Hanuschak-Keough strata mean weight, as defined on page 4, for the Hanuschak-Keough strata mean estimate

= the Hanuschak-Keough strata median weight, as defined on page 4, for the Hanuschak-Keough strata median estimate

= the modified weight, as defined on page 5, for the modified weighted estimate

= 1 if a Resident Farm Operator (RFO), and
= 0 if otherwise
for the open estimate

z_{mlkj} = peak number of hired workers for tract m, within segment j, substratum k and land use stratum l

The variance of the estimated peak number of hired workers is defined as follows:

$$var(\hat{Y}_{STATE}) = \sum_{l=1}^{s} \sum_{k=1}^{p_l} \sum_{j=1}^{r_{lk}} \frac{\bar{e}_{lk.} - 1}{\bar{e}_{lk.}} \frac{r_{lk}}{r_{lk} - 1} (\hat{Y}'_{lkj} - \tilde{Y}'_{lk.})^2$$

where

$$\tilde{Y}'_{lk.} = \sum_{j=1}^{r_{lk}} \frac{\hat{Y}'_{lkj}}{r_{lk}}$$

$$\bar{e}_{lk.} = \sum_{j=1}^{r_{lk}} \frac{e_{lkj}}{r_{lk}}$$

When computing the 17 regional level estimates and variances, the estimates and variances for each state within a given region were summed together.

Appendix E: Univariate paired t-tests

Univariate paired t-tests were conducted at both the regional and the state level for the peak number of hired workers. A t statistic was generated for the total difference (both at the regional and the state level).

Using the statistics defined in Appendix D, the difference in the expanded estimates was initially calculated at the segment level:

$$d_{lkj} = e_{lkj}\hat{Y}_{lkj,alternative} - e_{lkj}\hat{Y}_{lkj,open}$$

$$= e_{lkj}\hat{Y}_{lkj,difference}$$

$$= \hat{Y}'_{lkj,difference}$$

And the difference in the state level expanded estimates was:

$$\hat{Y}_D = \sum_{l=1}^{s}\sum_{k=1}^{p_l}\sum_{j=1}^{r_{lk}} d_{lkj}$$

For regional level differences, $\hat{Y}_D$ was summed to the regional level for each state within the given region.

The variances of the differences was also calculated at the state and regional levels. We used the same formula as in Appendix D, except that $\hat{Y}'_{lkj,difference}$ was substituted for $\hat{Y}'_{lkj}$.

The univariate paired t-test:

$Y_D = Y_{alternative} - Y_{open}$

To test: H_0: $Y_D = 0$
H_A: $Y_D <> 0$,

where

$$t = \frac{\hat{Y}_D}{s.e.\ (\hat{Y}_D)}$$

and reject H_O *if* $|t| > t_{alpha}$

Z tables were used in obtaining significance levels since the t is approximated by the z when sample sizes are large. Therefore, $t_{alpha} = t_{.05} = 1.96$.